* Is it accurate enough for your needs? Does it generalize well?
* Does it perform better than “the obvious guess”? Better than whatever estimate you currently use?
* Do the results of the model (coefficients, clusters, rules) make sense in the context of the problem domain?
* Confusion Matrix, Trade-off between precision & recall.

**Model Evaluation**

As the dataset is perfectly balanced, we fixed the accuracy threshold to be 0.5.

**Model Validation**

**Attributes: Age, Gender, Height, Weight, Systole, Diastole, Cholesterol, Glucose, Smoking, Alcohol, Active, BMI, BloodPressure**

**Decision Tree**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model variable** | **Attributes included** | **Attribute reason** | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1-score** |
| tree\_fit | Glucose  Smoking  Active  Alcohol | Musabbir’s default attribute in code | 54.17476 | 60.03394 | 17.6324 | 0.2725872 |
| tree\_fit2 | Age  BMI  Glucose  Active  BloodPressure | Shawn’s feature selection | 70.67961 | 73.41939 | 62.3676 | 0.6744574 |
| tree\_fit3 | ALL | Just to try out | 72.22694 | 75.18995 | 64.12461 | 0.6921784 |
| tree\_fit4 | Age  Gender  Cholesterol  Glucose  Smoking  Alcohol  Active  BMI  BloodPressure | BMI replace height and weight.  BloodPressure replace Systole and Diastole | 71.28034 | 72.95036 | 65.19626 | 0.688557 |

- Moved BloodPressure attribute inclusion into variable data at the top section, because before that testing\_data and training\_data missing attribute BloodPressure.

- Included outlier filtering for Height.

- Recalculate BMI and BloodPressure after filtering outliers.

**CHANGED**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model variable** | **Attributes included** | **Attribute reason** | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1-score** |
| tree\_fit | Systol  Cholesterol  Age  Diastole  Weight | Musabbir’s default attribute in code | 72.35528 | 75.45061 | 64.34061 | 0.6945412 |
| tree\_fit2 | Age  BMI  Glucose  Active  BloodPressure | Shawn’s feature selection | 71.46933 | 74.4577 | 63.31079 | 0.6843329 |
| tree\_fit3 | ALL | Just to try out | 72.35528 | 75.45061 | 64.34061 | 0.6945412 |
| tree\_fit4 | Age  Gender  Cholesterol  Glucose  Smoking  Alcohol  Active  BMI  BloodPressure | BMI replace height and weight.  BloodPressure replace Systole and Diastole | 72.05496 | 74.01656 | 65.93913 | 0.6974476 |

**Naïve Bayes**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model variable** | **Attributes included** | **Attribute reason** | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1-score** |
| nb\_model | Glucose  Smoking  Active  Alcohol | Musabbir’s default attribute in code | 54.55704 | 55.52805 | 33.54517 | 0.4182397 |
| nb\_model2 | Age  BMI  Glucose  Active  BloodPressure | Shawn’s feature selection | 70.86165 | 74.48716 | 61.08411 | 0.671231 |
| nb\_model3 | ALL | Just to try out | 71.37136 | 75.98617 | 60.24922 | 0.6720879 |
| nb\_model4 | Age  Gender  Cholesterol  Glucose  Smoking  Alcohol  Active  BMI  BloodPressure | BMI replace height and weight.  BloodPressure replace Systole and Diastole | 71.34709 | 74.00087 | 63.45171 | 0.6832148 |

- Moved BloodPressure attribute inclusion into variable data at the top section, because before that testing\_data and training\_data missing attribute BloodPressure.

- Included outlier filtering for Height.

- Recalculate BMI and BloodPressure after filtering outliers.

**Naïve Bayes**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model variable** | **Attributes included** | **Attribute reason** | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1-score** |
| nb\_model | Glucose  Smoking  Active  Alcohol | Musabbir’s default attribute in code | 54.29837 | 55.23881 | 33.95327 | 0.4205616 |
| nb\_model2 | Age  BMI  Glucose  Active  BloodPressure | Shawn’s feature selection | 71.06389 | 74.80359 | 61.46634 | 0.6748228 |
| nb\_model3 | ALL | Just to try out | 72.00991 | 76.84577 | 61.11282 | 0.6808219 |
| nb\_model4 | Age  Gender  Cholesterol  Glucose  Smoking  Alcohol  Active  BMI  BloodPressure | BMI replace height and weight.  BloodPressure replace Systole and Diastole | 71.76214 | 74.56596 | 64.0332 | 0.6889936 |